A revised δf algorithm for nonlinear PIC simulation

S. J. Allfrey and R. Hatzky†

simon.allfrey@epfl.ch

PPB 114,
Centre de Recherches en Physique des Plasmas,
Association EURATOM-Confédération Suisse,
Ecole Polytechnique Fédérale de Lausanne,
1015 Lausanne, Switzerland.

†*Computer Centre of the Max-Planck-Gesellschaft*
and the Institute for Plasma Physics
D-85748 Garching, Germany.

Abstract

An equation for the evolution of δf is shown to be redundant in the δf particle in cell (PIC) simulation scheme. Having eliminated this equation, an adaptive f_0 construction is shown now to follow intuitively and to be straight forward to implement.

1 Introduction

The particle in cell (PIC) method provides a means to computationally simulate systems described by kinetic equations. The positions of a large number of 'marker particles' (markers) are evolved along the characteristics of the underlying kinetic equations. Information from the simulation (including that required in the calculation of the characteristics) is obtained through statistical approximation of moments of the distribution function. The reliability of such simulations thus depends on the accuracy of the integration along the characteristics and the statistical error in the approximation of the moments. These are the two essential aspects of the PIC method, the issues regarding the second are the subject of this paper.

The authors come from the background of gyrokinetic [1] PIC simulation, and it is from this perspective that we shall discuss these issues. In this field it has long been appreciated that the statistical noise associated with the PIC method can be considerably reduced by using the δf method [2], 'simulating only the perturbation to the distribution function'. To describe the δf in this way is rather misleading as it does not clearly differentiate between the two aspects of the technique given above. An excellent explanation of how the δf algorithm *does* work was given by Aydemir [3], who in making the connection with Monte Carlo methods, showed δf to be an example of the control variates method of variance reduction. Control variates is not the only variance reduction technique available under a Monte Carlo scheme, importance sampling can also yield large benefits in terms of the reduction of the error in moment estimates. Recently Hatzky *et al.* [4] showed an example of the two methods used in conjunction, resulting in dramatic improvements in the quality of gyrokinetic simulations.

Gyrokinetic PIC codes have generally been applied in the first instance to linearised problems [5] which *do* require an equation for the evolution of δf . It is perhaps because of this that it was overlooked that such an equation is redundant for nonlinear calculations. The elimination of this equation not only results in a more elegant, efficient and more easily understood method, but also clears the way for effective implementation of an 'adaptive f_0' method.

2 The statistical error in the δf method

Readers unfamiliar with the techniques and terminology of the Monte Carlo method are referred to Hammersley & Handscomb [6] chapters 2-5. Fishman [7] also provides an extensive reference.

We consider a phase space volume preserving (symplectic) kinetic equation of the form

$$\frac{\mathrm{d}}{\mathrm{d}t} f \doteq \frac{\partial}{\partial t} f + \dot{\mathbf{Z}} \cdot \frac{\partial}{\partial \mathbf{Z}} f = 0, \tag{1}$$

where $f(\mathbf{Z}\,, t)$ is the distribution function of some population of N_s particles,

$$\int f(\mathbf{Z}\,)\ \mathrm{d}\Gamma = N_s, \qquad \mathrm{d}\Gamma = \mathcal{J}\ \mathrm{d}\mathbf{Z}\,, \tag{2}$$

where $\mathcal{J}$ is the phase space Jacobian, $\mathbf{Z}$ are the phase space coordinates and $\dot{\mathbf{Z}}$ ($\doteq \mathrm{d}\mathbf{Z}\,/\mathrm{dt}$) describe the characteristics. We postpone discussion of the collisional case to section 5.

The point of departure for the simulation is the choice of a (large) number N of points in phase space randomly chosen according to some probability density function (PDF) $p(\mathbf{Z}\,)$, the sampling distribution, satisfying

$$\int p(\mathbf{Z}\,)\ \mathrm{d}\Gamma = 1. \tag{3}$$

The simulation proceeds by evolving the position of these marker points or markers. Since the equations of motion describing the characteristics are symplectic we have

$$\frac{\mathrm{d}}{\mathrm{d}t} p = 0. \tag{4}$$

The probability density of both the particles described by f and of the markers is constant along a characteristic. Calculation of the characteristics will typically require evaluation of various moments of f.

To give a concrete example, the particles represented by f may respond to a field ϕ obeying the equation

$$L\phi(\mathbf{x}\,) = \rho(\mathbf{x}\,) \doteq \int f(\mathbf{Z}\,)\ \mathcal{J}_{\mathbf{v}}\ \mathrm{d}\mathbf{v} \tag{5}$$

for some operator L, with $\mathcal{J}_{\mathbf{v}}$ the velocity space Jacobian. The integral on the right hand side is taken over the velocity coordinates of phase space, thus representing a particle number density. Resolving this equation using the finite element

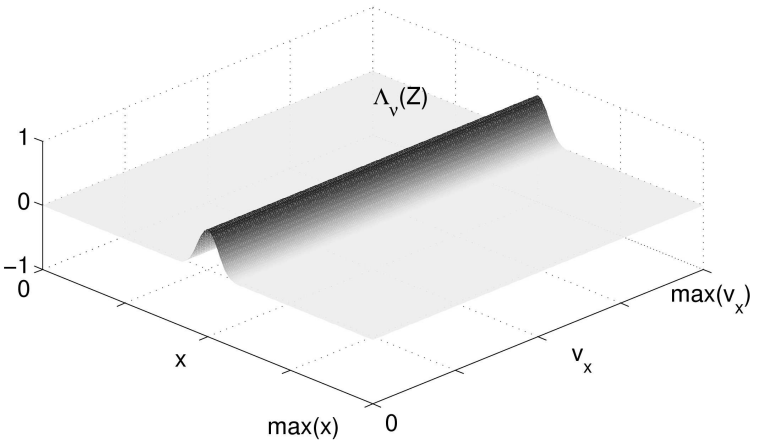

Figure 1: A typical finite element Λ_ν, projected into (x, v_x) space. Λ_ν is independent of v_x and localised in x, leading to an inevitable variance in approximation to moments associated with it.

method [8] yields

$$\phi \doteq \phi_\nu \Lambda_\nu , \tag{6}$$

$$A_{\nu\nu'}\phi_{\nu'} = b_\nu , \tag{7}$$

$$A_{\nu\nu'} \doteq (L\Lambda_{\nu'}, \Lambda_\nu), \qquad b_\nu \doteq (\rho, \Lambda_\nu). \tag{8}$$

Here summation convention over repeated indicies is assumed and the inner product is defined over configuration space as

$$(A, B) \doteq \int A(\mathbf{x})B(\mathbf{x})\ \mathcal{J}_\mathbf{x}\ \mathrm{d}\mathbf{x}\,. \tag{9}$$

Where $\mathcal{J}_\mathbf{x}$ is the configuration space Jacobian. By defining the finite elements Λ_ν to be functions of $\mathbf{Z}$ independent of the velocity coordinates (*i.e.* is functionally dependent only on $\mathbf{x}$, see figure 1) we can write the required moments of the distribution function as

$$b_\nu = \int f(\mathbf{Z})\Lambda_\nu(\mathbf{Z})\ \mathrm{d}\Gamma \tag{10}$$

Nonlinearly it has been found that even small errors in the evaluation of these moments can cause a systematic corruption of the simulation in a relatively short period of time. Therefore our goal is to choose an estimator for b_ν, using the information we have at the marker position, that reduces as far as reasonably possible (*i.e.* up to computational limits) the error ϵ_{b_ν} in the estimate for b_ν and thus the

error in ϕ_ν given by

$$\epsilon^2_{\phi_\nu} = \left(\frac{\partial \phi_\nu}{\partial b_{\nu'}}\right)^2 \epsilon^2_{b_{\nu'}} = \left(A^{-1}_{\nu\nu'}\right)^2 \epsilon^2_{b_{\nu'}}. \tag{11}$$

As is clearly described by Aydemir [3], the crude Monte Carlo estimator for b_ν is given by writing b_ν as the expectation value of a function g_ν. The standard error ϵ_{b_ν} in the estimate is proportional to the square root of the variance σ_ν in this estimating function.

$$b_\nu = \langle g_\nu \rangle_p \simeq \frac{1}{N}\sum_{i=1}^{N} g_{\nu i}, \qquad \epsilon_{b_\nu} = \frac{\sigma_\nu}{\sqrt{N}}, \tag{12}$$

$$g_{\nu i} = g_\nu(\mathbf{Z}_i), \qquad g_\nu(\mathbf{Z}) = \frac{f(\mathbf{Z})\,\Lambda_\nu(\mathbf{Z})}{p(\mathbf{Z})}, \tag{13}$$

$$\langle g_\nu \rangle_p = \int g_\nu(\mathbf{Z}) p(\mathbf{Z})\, \mathrm{d}\Gamma, \tag{14}$$

$$\sigma^2_\nu = \int [\, g_\nu(\mathbf{Z}) - \langle g_\nu \rangle_p]^2 p(\mathbf{Z})\, \mathrm{d}\Gamma. \tag{15}$$

Here $\mathbf{Z}_i$ are the coordinates of the ith of the N markers. Note that the expression for ϵ_{b_ν}, the standard error in the estimate of b_ν is exact. Turning to the (unbiased) estimate of the variance σ^2_ν, the standard error $\epsilon_{\sigma^2_\nu}$ for the expression

$$\sigma^2_\nu \simeq \frac{1}{N-1}\sum_{i=1}^{N}(g_{\nu i} - \langle g_\nu \rangle_p)^2, \qquad \epsilon_{\sigma^2_\nu} = \frac{\sigma^2_\nu}{\sqrt{(N-1)/2}}, \tag{16}$$

is exact only for normally distributed $g_{\nu i}$. The key to understanding statistical optimisation (noise reduction) in the PIC method is a clear insight into Eq. (15).

Let us first consider the technique of importance sampling [9]. In this scheme one chooses $p(\mathbf{Z}) = N_s^{-1} f(\mathbf{Z})$, leading to

$$\sigma^2_\nu = N_s \int [\, \Lambda_\nu(\mathbf{Z}) - \langle \Lambda_\nu \rangle_p]^2 f(\mathbf{Z})\, \mathrm{d}\Gamma, \tag{17}$$

$$\langle \Lambda_\nu \rangle_p = \int \Lambda_\nu(\mathbf{Z}) p(\mathbf{Z})\, \mathrm{d}\Gamma.$$

Although this scheme has zero variance (error) in the estimation of the total number of particles (consider $\Lambda_\nu(\mathbf{Z}) = 1$), spatially localised Λ_ν, used in representing

e.g. $\phi(\mathbf{x})$, will inevitably give rise to a non zero variance. Furthermore the variance will increase with the resolution (decrease in volume) of the finite elements. Thus regularly spaced finite elements in a coordinate system with a strongly varying Jacobian (*e.g.* cylindrical coordinates) will favor accuracy in the estimates relating to the larger elements (those at large r) at the expense of the smaller (those near the axis) if the sampling distribution $p(\mathbf{Z})$ is not chosen to accommodate for this effect. Choosing the finite elements to be of equal volume also removes this effect.

Now suppose that we know some readily calculable function $f_0(\mathbf{Z})$ that is close to $f(\mathbf{Z})$ in the sense that everywhere $|f - f_0| \ll f$. We can use the extra information represented by the knowledge of a suitable f_0 to improve our estimate for b_ν as follows. Write

$$b_\nu = \int f_0(\mathbf{Z})\Lambda_\nu(\mathbf{Z})\, \mathrm{d}\Gamma + \int [f(\mathbf{Z}) - f_0(\mathbf{Z})]\Lambda_\nu(\mathbf{Z})\, \mathrm{d}\Gamma. \tag{18}$$

Now it is assumed we can calculate the first integral with no statistical error and need only statistically approximate the second. The standard error in *this* approximation will be proportional to the square root of the variance of the function

$$h_\nu(\mathbf{Z}) = \frac{[f(\mathbf{Z}) - f_0(\mathbf{Z})]\Lambda_\nu(\mathbf{Z})}{p(\mathbf{Z})}. \tag{19}$$

We gain a reduction in the error of the order

$$\frac{\sigma_{h_\nu}}{\sigma_{g_\nu}} \sim \frac{|f - f_0|}{f}. \tag{20}$$

This is the control variates method of variance reduction known in the gyrokinetic community as the δf method with $\delta f \doteq f - f_0$. A discussion of the error under a scheme involving combined control variates and importance sampling is given in the appendix.

3 Equations for δf

The point that we now wish to stress is that the δf construction is conceptually separate from the integration of the characteristics and in particular there is no requirement for an additional 'equation for the evolution of δf'. f is a constant

along the trajectories and the (in some sense arbitrary) function f_0 is chosen to be calculable at any point in phase space. Thus δf (actually h_ν Eq. (19)) may be evaluated using

$$\delta f\left(\mathbf{Z}_i(t)\right) = f\left(\mathbf{Z}_i(t_0)\right) - f_0(\mathbf{Z}_i(t)) \tag{21}$$

for any marker i as and when it is required in the calculation of the estimates.

The validity and utility of the δf Ansatz remains untouched by the foregoing, *e.g.* physics motivated decompositions such as

$$f = f_{\rm eq} + \delta f, \tag{22}$$

$$f = f_{\rm ss} + \delta f. \tag{23}$$

where in the first case δf represents the departure from some initial equilibrium state and in the second a fluctuation of zero ensemble average about some steady state. Furthermore linearised simulation *does* require an equation for the evolution of δf . In the linear scheme the markers are evolved along the equilibrium characteristics, but f is evolved as if the nonlinear trajectory passing through the marker's coordinates at time t had been followed. In both the linearised and nonlinear cases δf will obey

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta f = -\frac{\mathrm{d}}{\mathrm{d}t}f_0 = -\frac{\partial}{\partial t}f_0 - \dot{\mathbf{Z}}\cdot\frac{\partial}{\partial \mathbf{Z}}f_0. \tag{24}$$

Typically f_0 will be time independent. In the case of nonlinear simulation where the markers follow the characteristics and f_0 is independent of time the solution of Eq. (24) is just Eq. (21). If Eq. (22) is chosen and we decompose the trajectories into equilibrium and perturbation parts *i.e.*

$$\dot{\mathbf{Z}} = \dot{\mathbf{Z}}\,|_{\rm eq} + \dot{\mathbf{Z}}\,|_{\rm pert}. \tag{25}$$

We have

$$\frac{\partial}{\partial t}f_{\rm eq} = \dot{\mathbf{Z}}\,|_{\rm eq}\cdot\frac{\partial}{\partial \mathbf{Z}}f_{\rm eq} = 0, \tag{26}$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\delta f = -\dot{\mathbf{Z}}\,|_{\rm pert}\cdot\frac{\partial}{\partial \mathbf{Z}}f_{\rm eq}, \tag{27}$$

the usual form of the $\dot{\delta f}$ equation. Thus the quantities

$$w_i \doteq \frac{1}{N}\frac{\delta f\left(\mathbf{Z}_i\right)}{p(\mathbf{Z}_i)}, \tag{28}$$

referred to as the marker weights, can be calculated directly using Eq. (21). Figure (2) shows a comparison of PIC simulations with and without this equation for $\dot{\delta f}$ used.

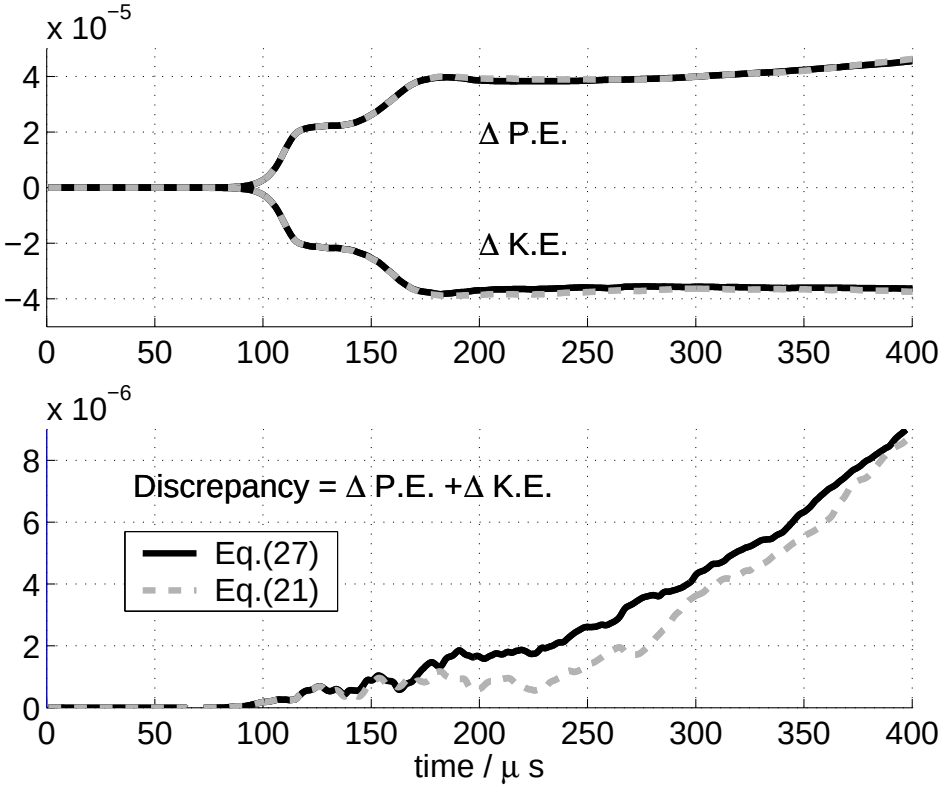

Figure 2: Comparison of nonlinear gyrokinetic turbulence simulations in a theta-pinch i) integrating Eq. (27) solid line and ii) using Eq. (21) dashed line. The upper figure shows the time evolution of the change in the potential and kinetic energy. The lower figure shows the departure from energy conservation. The results are very similar, with a slight improvement in energy conservation under the new scheme. Energy is normalised to the total equilibrium thermal energy of the system.

4 Implications for adaptive f_0

The control variates method uses 'external' information (information not currently carried by the markers), knowledge of an appropriate f_0, to reduce the error in a statistical approximation. Clearly, at early times during a simulation the choice of $f_0 = f_{\rm eq}$ is indicated. But as f evolves away from $f_{\rm eq}$, use of the latter will

no longer provide a small variance estimator *i.e.* $\sigma_\nu^2(t)$ will increase with time. If possible it would be desirable to evolve f_0 in such a way as to 'follow' f.

One such technique has been implemented in collisional Monte Carlo simulations [10]. In this case an appropriate choice for $f_0(t)$ was a shifted Maxwellian distribution, evolved using fluid equations.

Here we discuss a more general technique. Choose for f_0 some approximation to f at the previous time step *i.e.* seek to approximate the following scheme

$$f(t^0) = f_0(t^0) = f_{\rm eq} \quad \Rightarrow \delta f\,(t^0) = 0 \tag{29}$$

$$\begin{aligned} f_0(t^n) &= f_0(t^{n-1}) + \delta f\,(t^{n-1}) \\ &= f_{\rm eq} + \sum_{j=0}^{n-1} \delta f\,(t^j) \end{aligned} \tag{30}$$

in which δf represents only the change in f between time-steps. One approach to this approximation is to use

$$f_0(\mathbf{Z}\,, t^n) = f_{\rm eq}(\mathbf{Z}\,) + \sum_i \sum_{j=1}^{n} \beta_i(t^j) B_i(\mathbf{Z}\,), \tag{31}$$

where the $B_i(\mathbf{Z}\,)$ are a general set of functions covering the phase space. We note that having eliminated Eq. (24) there are no longer any requirements of continuity on f_0, we require only that we be able to evaluate it at any point in phase space. While f_0 changes between time steps it is constant in time during the time integration. For the case of orthogonal piecewise constant functions we do not have to explicitly solve for the coefficients $\beta_i(t^j)$, they can be calculated directly as

$$\beta_i(t^j) = \frac{\int \delta f\,(\mathbf{Z}\,, t^{j-1}) B_i(\mathbf{Z}\,)\,{\rm d}\Gamma}{\int B_i^2(\mathbf{Z}\,)\,{\rm d}\Gamma}. \tag{32}$$

This technique can be seen as giving the δf method a memory. Structures persistent in time, *i.e.* existing over many integration time steps, will be multiply sampled and information about them 'layered down' according to Eqs. (30),(31).

We note that in practice the refinement of a useful B_i 'mesh' will be set by consideration of the statistical error in calculating Eq. (32). In those dimensions which exhibit little variance, the B_i could be chosen to be constant. With deeper insight

into the problem at hand, representations more sophisticated than Eq. (31) may be appropriate. For example structure may manifest itself in certain dimensions while in others only thermal dependence is seen, there fitting of 'local Maxwellians' in the thermal coordinates may be the best approach.

A fundamental principle is that as far as practical we should not discard any information we have available to us (*e.g.* from the previous time step) when making our statistical approximations.

In general the algorithm permits any choice of functional representation for f_0, and is in principle not critically sensitive to this choice. A poor choice for f_0 will not necessarily be catastrophic, but simply increase the error in the estimates. The nature of the problem at hand will set the tolerance to these errors.

5 Discussion

A δf PIC scheme derives its advantage over a 'full-f' scheme through a control variates reduction of statistical error in moment approximations. No integration of $\dot{\delta f}$ is required along the trajectories of the markers. The authors have implemented this $\dot{\delta f}$ free scheme as described in section 3. Such a scheme has several advantages over using a time evolution equation for δf (see Eq. (27)). The computational work load is reduced and no additional integration error is introduced. With no requirements of continuity on f_0, adaptation of this background function in time becomes a practical possibility. A further optimisation of the adaptive f_0 scheme can be achieved through combination with a scheme involving importance sampling [4]. Evaluation of an optimal adaptive f_0 scheme for collisionless gyrokinetic simulation is in progress and will be the subject of a forthcoming publication.

If collisions are to be included in the simulation an additional operator must included on the right hand side of Eq. (1)

$$\frac{\mathrm{d}}{\mathrm{d}t} f = C(f) \tag{33}$$

f is no longer conserved along a trajectory (the markers no longer move along characteristics) as particles are scattered into and out of the phase space locale of the marker. If a tractable form for $C(f)$ could be found, this scheme discussed in

this paper would remain essentially unchanged but for the addition of this equation for the evolution of f along a marker trajectory. However such a tractable form is usually not available since an accurate description must include second order velocity space gradients of the distribution function. The collision operator is most often simulated by introducing a suitable stochastic component to the trajectories of the markers [11]. The marker trajectories then mimic the microscopic (Langevin) trajectories of the particles underlying the distribution f. However the correspondence between the Langevin equations and the kinetic equations holds only in the ensemble average sense (average over all possible outcomes of collisions over all phase space). This is the origin of the additional phase space dimension introduced in [11]. Simulating the collision operator in this way introduces an additional statistical error for each marker both with respect to accurately simulating Eq. (33) and in any calculation of the change in p along the now non-symplectic trajectory. With this caveat we note that it is still possible to construct equations for $\dot{f}$ and $\dot{p}$ which are independent of f_0 and thus permit implementation of the adaptive f_0 scheme we have described.

Acknowledgments

This work was partly supported by the Swiss National Science Foundation. Our simulations were carried out on the IBM 'Regatta' of the Joint Computer Center of the Max-Planck Gesellschaft, Garching and the SGI Origin 3800 of the Ecole Polytechnique Fédérale de Lausanne. The authors gratefully acknowledge the support of both of these institutions.

Appendix: Error with combined control variates and importance sampling

Consider the Monte Carlo scheme described in section 2. The variance of our Monte Carlo estimator of a function $h(\mathbf{Z})$ using a probability distribution $p(\mathbf{Z})$ is given by

$$\sigma_{h_p}^2 = \int \left[\frac{h}{p} - \left\langle \frac{h}{p} \right\rangle_p\right]^2 p \; \mathrm{d}\Gamma, \tag{34}$$

$$\langle h \rangle_p \doteq \int hp \; \mathrm{d}\Gamma. \tag{35}$$

Thus

$$\begin{aligned}\sigma_{h_p}^2 &= \int \frac{h^2}{p} \; \mathrm{d}\Gamma - 2 \left\langle \frac{h}{p} \right\rangle_p \int h \; \mathrm{d}\Gamma + \left\langle \frac{h}{p} \right\rangle_p^2 \\ &= \int \frac{h^2}{p} \; \mathrm{d}\Gamma - \left[\int h \; \mathrm{d}\Gamma\right]^2. \end{aligned} \tag{36}$$

For positive h this expression obtains a minimum (zero) with respect to the function p when $p \propto h$ *i.e.*

$$p = \alpha \, h, \qquad\qquad \alpha^{-1} = \int h \; \mathrm{d}\Gamma. \tag{37}$$

We now compare the variance with a full-f importance sampling scheme and a δf optimised [4] scheme for the case of spatially localised functions Λ as discussed in section 2. In the first case we have

$$h = \Lambda f, \qquad\qquad p = N_s^{-1} f, \qquad\qquad N_s \doteq \int f \; \mathrm{d}\Gamma, \tag{38}$$

and in the second

$$h = \Lambda \delta f \, , \qquad\qquad p = \delta N_s^{-1} |\delta f \, |, \qquad\qquad \delta N_s \doteq \int |\delta f \, | \; \mathrm{d}\Gamma. \tag{39}$$

For the sake of clarity we have used a 'perfect' optimisation $|\delta f_{\mathrm{opt}}| = |\delta f \, |$. The variances are given by

$$\sigma_{\mathrm{imp}}^2 = N_s \int \Lambda^2 f \; \mathrm{d}\Gamma - \left[\int \Lambda f \; \mathrm{d}\Gamma\right]^2, \tag{40}$$

$$\sigma^2_{\delta f\,\mathrm{opt}} = \delta N_s \int \Lambda^2 |\delta f \mid \mathrm{d}\Gamma - \left[\int \Lambda \delta f \;\; \mathrm{d}\Gamma\right]^2 . \tag{41}$$

For localised Λ

$$\int \Lambda f \;\mathrm{d}\Gamma \ll \int f \;\mathrm{d}\Gamma = N_s, \tag{42}$$

$$\int \Lambda \delta f \;\; \mathrm{d}\Gamma \ll \int |\delta f \mid \mathrm{d}\Gamma = \delta N_s, \tag{43}$$

and the variances are dominated by the first term. Thus, due to the localisation of Λ, the ratio of the variances of the 'best' full-f and δf methods is given by

$$\frac{\sigma^2_{\delta f\,\mathrm{opt}}}{\sigma^2_{\mathrm{imp}}} \sim \left(\frac{\langle |\delta f \mid \rangle}{\langle f \rangle}\right)^2, \tag{44}$$

whilst the full-f importance sampling scheme has zero error for $\Lambda = 1$.

References

[1] W. W. Lee. Gyrokinetic approach in particle simulation. *Phys. Fluids*, 26:556, 1983.

[2] S. E. Parker and W. W. Lee. A fully nonlinear characteristic method for gyrokinetic simulation. *Phys. Fluids B*, 5:77, 1993.

[3] A. Y. Aydemir. A unified monte carlo interpretation of particle simulations and applications to non-neutral plasmas. *Phys. Plasmas*, 1:822, 1994.

[4] R. Hatzky, T. M. Tran, A. Könies, R. Kleiber, and S. J. Allfrey. Energy conservation in a nonlinear gyrokinetic particle-in-cell code for ion-temperature-gradient-driven modes in θ-pinch geometry. *Phys. Plasmas*, 31:898, 2002.

[5] M. Fivaz, S. Brunner, G. de Ridder, O. Sauter, T. M. Tran, J. Vaclavik, and L. Villard. Finite element approach to global gyrokinetic particle-in-cell simulations using magnetic coordinates. *Comput. Phys. Commun.*, 111:27, 1998.

[6] J. M. Hammersley and D. C. Handscomb. *Monte Carlo Methods*. Methuen and Co Ltd., London, 1964. Also Chapman and Hall, 1965.

[7] G. S. Fishman. *Monte Carlo: Concepts, Algorithms, and Applications*. Springer-Verlag, New York, 1999.

[8] G. Strang and G. J. Fix. *An analysis of the finite element method*. Prentice-Hall, Englewood Cliffs, New Jersey, 1973.

[9] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. *J. Chem Phys*, 21:1087, 1953.

[10] S. Brunner, E. Valeo, and J. A. Krommes. Collisional delta-f scheme with evolving background for transport times scale simulations. *Phys. Plasmas*, 6(12):4504, 1999.

[11] Y. Chen and R. White. Collisional delta-f method. *Phys. Plasmas*, 4:3591, 1997.